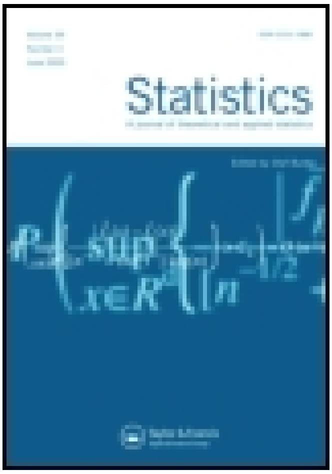

Statistics: A Journal of Theoretical and Applied Statistics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gsta20

A doubly skewed normal distribution

Barry C. Arnold[a], Héctor W. Gómez[b] & Hugo S. Salinas[c]

[a] Statistics Department, University of California, Riverside, CA, USA

[b] Departamento de Matemáticas, Facultad de Ciencias Básicas, Universidad de Antofagasta, Antofagasta, Chile

[c] Departamento de Matemática, Facultad de Ingeniería, Universidad de Atacama, Atacama, Chile

Published online: 29 May 2014.

To cite this article: Barry C. Arnold, Héctor W. Gómez & Hugo S. Salinas (2014): A doubly skewed normal distribution, Statistics: A Journal of Theoretical and Applied Statistics, DOI: 10.1080/02331888.2014.918618

To link to this article: http://dx.doi.org/10.1080/02331888.2014.918618

Statistics, 2014
http://dx.doi.org/10.1080/02331888.2014.918618

Taylor & Francis
Taylor & Francis Group

A doubly skewed normal distribution

Barry C. Arnold[a], Héctor W. Gómez[b*] and Hugo S. Salinas[c]

[a]*Statistics Department, University of California, Riverside, CA, USA;* [b]*Departamento de Matemáticas, Facultad de Ciencias Básicas, Universidad de Antofagasta, Antofagasta, Chile;* [c]*Departamento de Matemática, Facultad de Ingeniería, Universidad de Atacama, Atacama, Chile*

(*Received 2 January 2012; accepted 23 December 2013*)

We consider a distribution obtained by combining two well-known mechanisms for generating skewed distributions. In this manner we arrive at a flexible model which subsumes and extends several skew distributions which have been discussed in the literature. One approach to the problem of generating skewed distributions was first popularized by Azzalini [A class of distributions which includes the normal ones. Scand J Stat. 1985;12:171–178]. The single constraint skew normal distribution that was studied by Azzalini is of the form

$$f_{\mathrm{SN}}(x) = 2\phi(x)\Phi(\alpha x),$$

where ϕ and Φ denote, respectively, the standard normal density and distribution function and $\alpha \in \mathbb{R}$ is a skewing parameter. Multiple constraint variations of this distribution have also been considered. The second skewing approach that we will consider was proposed by Mudholkar and Hutson [The epsilon-skew-normal distribution for analyzing near-normal data. J Statist Plann Inference. 2000;83:291–309] and called an epsilon-skew-normal distribution. The combination of an Azzalini mechanism with that of Mudholkar and Hutson is investigated in this paper with special focus on the distributions obtained using the standard normal as the base distribution. The resulting flexible model includes both unimodal and bimodal cases and can be expected to fit a wider variety of data configurations than either of the models involving a single skewing mechanism. Distributional and inferential properties of the doubly skewed model are discussed and the model is used to obtain improved fits to two well-known data sets.

Keywords: epsilon-skew-normal; two-piece skew-normal; skew-normal distribution; doubly skewed

1. Introduction

Beginning with a base density $f(x)$ and a given distribution function $G(x)$, a skewed version of the density f of the Azzalini type [1] will be of the form

$$f_{\mathrm{A}}(x) \propto f(x)G(\alpha x). \tag{1}$$

Multiple constraint versions of this construction will involve k skewing distribution functions instead of just one and will be of the form

$$f_{\mathrm{MA}}(x) \propto f(x)\prod_{j=1}^{k} G_j(\psi_j(x)). \tag{2}$$

in which the G_j's are distribution functions and the ψ_j's are usually simple functions, often linear.

*Corresponding author. Email: hector.gomez@uantof.cl

Examples of this genre which have appeared in the literature include:

(i) The multiple linear constraint normal model [2]:

$$f(x) \propto \phi(x) \prod_{j=1}^{k} \Phi(\alpha_{0j} + \alpha_{1j}x) \tag{3}$$

in which and throughout the rest of this paper ϕ and Φ denote, respectively, the standard normal density and distribution function.

(ii) The two piece normal (TN) distribution [3]:

$$f_{\mathrm{TN}}(x) = c_\alpha \phi(x)\Phi(\alpha|x|). \tag{4}$$

(iii) The extended two piece normal (ETN) distribution [4]:

$$f_{\mathrm{ETN}}(x) = 2c_\alpha \phi(x)\Phi(\alpha|x|)\Phi(\beta x). \tag{5}$$

In (ii) and (iii), the normalizing constant (to guarantee the density integrates to 1) is explicitly available. Specifically

$$c_\alpha = \frac{2\pi}{\pi + 2\arctan(\alpha)}. \tag{6}$$

This is unusual. More typically, in multiple constraint models, the normalizing constant must be evaluated by numerical integration for any specific set of parameter values.

An alternative skewing mechanism was proposed by Mudholkar and Hutson.[5] Using their approach, the epsilon skewed version of a base density $f(x)$ is of the form:

$$f_\varepsilon(x) = \frac{f(x/(1+\varepsilon))I(x<0) + f(x/(1-\varepsilon))I(x \geq 0)}{1-\varepsilon(1-2F(0))}. \tag{7}$$

where $-1 < \varepsilon < 1$ is the skewing parameter, $I(B)$ is the indicator function of the set B, and F is the distribution function corresponding to the density f. If, as in the normal case studied by Mudholkar and Hutson, the base density is symmetric, or less stringently if $F(0) = 1/2$, then the denominator in Equation (7) reduces to 1 and can be eliminated. In particular, if $f = \phi$ we have the epsilon-skew normal (ESN) density

$$f_{\mathrm{ESN}}(x;\varepsilon) = \phi\left(\frac{x}{1+\varepsilon}\right) I(x<0) + \phi\left(\frac{x}{1-\varepsilon}\right) I(x \geq 0). \tag{8}$$

This is the specific model introduced by Mudholkar and Hutson. Subsequently Arellano-Valle et al. [6] discuss the more general case in which the base density is an arbitrary symmetric density. In particular, they focused on the case in which the base density was a symmetric exponential power density (i.e. $f(x) \propto \exp(|x|^\gamma)$.

To indicate that a random variable X has density f_{ESN} we will write $X \sim \mathrm{ESN}(\varepsilon)$. Likewise, if we write $X \sim A(\alpha)$ we mean that X has f_A (defined in Equation (1)) as its density. The same convention is used for other subscripted densities introduced in the paper. To indicate that a base density f is symmetric about 0, we will append a subscript 0, thus f_0 will denote a generic density that is symmetric about 0. Likewise, to indicate that a distribution function G is symmetric (i.e. that $G(x) + G(-x) = 1$), we append a subscript 0, thus G_0.

The focus of the present paper is the study of the distribution obtained by applying both kinds of skewing mechanisms (the multiple constraint mechanism and the epsilon skewing mechanism) to the same base density. In particular we study the epsilon skewed version of the ETN density. Thus our base density is standard normal. The use of alternative base densities can be envisioned, but there are computational advantages associated with the normal density leading to more attractive expressions; for example the normalizing constant for the doubly skewed normal density is available in explicit form. Discussion of models obtained by applying more than one skewing mechanism to a non-normal base density will be deferred to a separate report.

The resulting distribution can be viewed as an extension of the epsilon skew normal model. In an earlier report Arellano-Valle et al. [7] discussed a different extension of the epsilon skew normal model. To avoid the possible confusion associated with two distinct extended epsilon skew normal models, and to emphasize the fact that the present models involve two skewing mechanisms, we will speak of them as being doubly skewed normal models (rather than extended epsilon skew normal models). As will be seen, the doubly skewed models add considerable flexibility since they include bimodal as well as unimodal cases.

The paper is organized as follows. Section 2 includes general discussion of double skewing. In Section 3, we develop the doubly skewed normal model and provide graphs of representative densities together with discussion of basic properties and the relationship of the doubly skewed normal model to other skew normal models which it includes and extends. In Section 4, distributional properties of a restricted but still flexible sub-model are investigated in detail. For this sub-model maximum likelihood estimation (MLE) is discussed and the corresponding Fisher information matrix is derived in Section 5. Application of the sub-model to two particular data sets is the subject of Section 6. Some details have been deferred to the appendix.

2. Doubly skewed models

Begin with a base density function f. The corresponding multiple constraint Azzalini-type skewed density is of the form:

$$f_{\mathrm{MA}}(x;\underline{\theta}) = C(\underline{\theta})f(x)\prod_{j=1}^{k} G_j(\psi_j(x);\theta_j). \tag{9}$$

where

$$C(\underline{\theta})^{-1} = \int_{-\infty}^{\infty} f(x)\prod_{j=1}^{k} G_j(\psi_j(x);\theta_j)\,\mathrm{d}x.$$

Typically the normalizing constant $C(\underline{\theta})$ will need to be evaluated by numerical integration, although in some special cases it can be evaluated analytically. The corresponding distribution function will be denoted by $F_{\mathrm{MA}}(x;\underline{\theta})$.

The corresponding doubly skewed model, obtained by applying ε-skewing to the skewed model (9) is then of the form

$$f_{\mathrm{DS}}(x;\varepsilon,\underline{\theta}) = C(\underline{\theta})\left[\frac{f_{\mathrm{MA}}(x/(1+\varepsilon);\underline{\theta})I(x<0) + f_{\mathrm{MA}}(x/(1-\varepsilon);\underline{\theta})I(x\geq 0)}{1-\varepsilon(1-2F_{\mathrm{MA}}(0;\underline{\theta}))}\right]. \tag{10}$$

Considerable simplification of this expression is possible in the single constraint case when $\psi_1(x) = x$ and both f and G are symmetric. Thus, using α as the Azzalini skewing parameter, the doubly skewed model is given by

$$f_{\mathrm{DS}}(x;\varepsilon,\alpha) = 2\frac{\begin{array}{c} f_0(x/(1+\varepsilon))G_0(\alpha(x/(1+\varepsilon)))\,I(x<0) \\ +f_0(x/(1-\varepsilon))G_0(\alpha(x/(1-\varepsilon)))I(x\geq 0)\end{array}}{1-\varepsilon(1-2F_0(0;\alpha))}. \tag{11}$$

where $F_0(0;\alpha) = 2\int_{-\infty}^{0} f_0(x)G_0(\alpha x)\,\mathrm{d}x$.

As will be described in the next section, if $f = \phi$ and $G = \Phi$ it is sometimes possible to identify precisely the form of the normalizing constant.

3. The doubly skewed normal distribution

The purpose of the present section is to introduce what we call the doubly skewed normal density which can be viewed as a uni/bimodal extension of the epsilon-skew-normal density obtained by replacing the density to be epsilon-skewed by an already skewed density. Such an extension is potentially relevant for practical applications, because there are far fewer distributions available for dealing with bimodal data than in the unimodal case.

3.1. *The family of doubly skewed normal densities*

The base density will be standard normal. After applying a particular two constraint Azzalini-type skewing mechanism, we arrive at the extended two piece normal distribution introduced by Arnold et al.[4]

$$f_{\mathrm{ETN}}(x) = 2c_\alpha \phi(x)\Phi(\alpha|x|)\Phi(\beta x).$$

In this case the normalizing constant, c_α, is explicitly available in the form

$$c_\alpha = \frac{2\pi}{\pi + 2\arctan(\alpha)}.$$

Upon applying the ε-skewing mechanism, the following density, called the doubly skewed normal (DSN) density, is obtained.

$$f_{\mathrm{DSN}}(x;\varepsilon,\alpha,\beta) = 2c_\alpha \left[\frac{\begin{array}{c}\phi(x/(1+\varepsilon))\Phi(\alpha|x/(1+\varepsilon)|)\Phi(\beta x/(1+\varepsilon))I(x<0)\\ +\phi(x/(1-\varepsilon))\Phi(\alpha|x/(1-\varepsilon)|)\Phi(\beta x/(1-\varepsilon))I(x\geq 0)\end{array}}{1-\varepsilon(1-2d_{\alpha,\beta})} \right]. \tag{12}$$

where

$$d_{\alpha,\beta} = F(0;\alpha,\beta) = \int_{-\infty}^{0} 2c_\alpha \phi(x)\Phi(\alpha|x|)\Phi(\beta x)\,\mathrm{d}x.$$

There is not an available analytic expression for the quantity $d_{\alpha,\beta}$ and it will be necessary to evaluate it numerically for given values of α and β. This will create some inconvenience when, for example, maximum likelihood estimates are to be computed; but not insurmountable inconveniences. The shapes of the density (12) for different parameter values are displayed in Figure 1.

There exists a possibility of singularity of the Fisher information matrix with the DSN model. Such is the case for certain well-known submodels, such as the Azzalini skew-normal model. One possible approach to avoid singularity is to consider power-models as in [8]. The possibility exists of extending the results to more robust models (see, for example, Gómez et al. [9]). This will be the subject of a separate report.

In the special case in which $\alpha = 0$, in which case we are dealing with an ε-skewed version of the basic Azzalini skew normal density, it is possible to evaluate the required normalizing

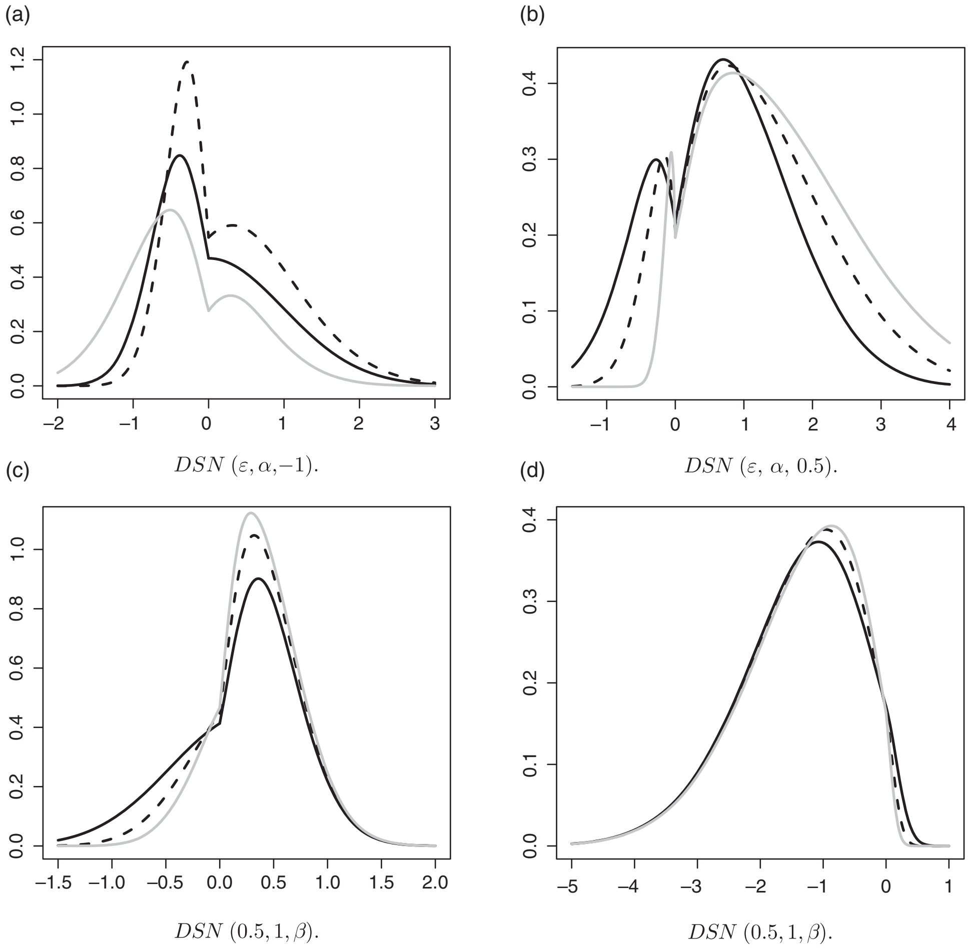

Figure 1. Examples of the DSN density for: (a) $(\varepsilon, \alpha) = (-0.5, 1)$ (black line), $(\varepsilon, \alpha) = (-0.6, 2)$ (dashed line) and $(\varepsilon, \alpha) = (-0.2, 3)$ (grey line); (b) $(\varepsilon, \alpha) = (-0.2, 3)$ (black line), $(\varepsilon, \alpha) = (-0.5, 4)$ (dashed line) and $(\varepsilon, \alpha) = (-0.8, 5)$ (grey line); (c) $\beta = 2$ (black line), $\beta = 3$ (dashed line) and $\beta = 4$ (grey line); (d) $\beta = -2$ (black line), $\beta = -3$ (dashed line) and $\beta = -4$ (grey line).

constant. We have

$$d_{0,\beta} = F(0; 0, \beta) = \frac{1}{2} - \frac{1}{\pi} \arctan \beta. \tag{13}$$

Thus the density will be

$$f_{\mathrm{DSN}}(x; \varepsilon, 0, \beta) = \frac{\phi(x/(1+\varepsilon))\Phi(\beta x/(1+\varepsilon))I(x<0) + \phi(x/(1-\varepsilon))\Phi(\beta x/(1-\varepsilon))I(x \geq 0)}{1/2 - (\varepsilon/\pi)\arctan \beta}. \tag{14}$$

This model is similar to, though not the same as, the model introduced by Gómez et al.[10] Figure 2 shows the shapes of the density (14) for different parameter values.

3.2. *Distributional properties of the doubly skewed model*

It is clear that the doubly skewed density (12) is continuous at 0 for every α, β and ε. However, it is not differentiable at 0 unless both α and β are equal to 0. The density can be unimodal or

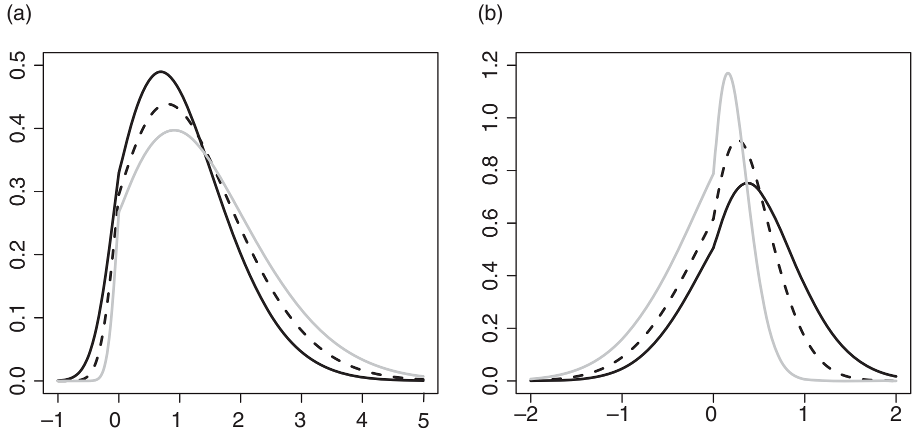

Figure 2. Examples of the DSN$(\varepsilon, 0, 2)$ density for: (a) $\varepsilon = -0.3$ (black line), $\varepsilon = -0.5$ (dashed line) and $\varepsilon = -0.7$ (grey line) and (b) $\varepsilon = 0.3$ (black line), $\varepsilon = 0.5$ (dashed line) and $\varepsilon = 0.7$ (grey line).

bimodal. It includes several previously studied models as special cases, as noted in the list of properties below.

Property 3.1 *The* DSN$(0, 0, 0)$ *density is the* $N(0, 1)$ *density.*

Property 3.2 *The* DSN$(\varepsilon, 0, 0)$ *density is the* ESN(ε) *density.*

Property 3.3 *The* DSN$(0, 0, \beta)$ *density is the* SN(β) *density.*

Property 3.4 *The* DSN$(0, \alpha, 0)$ *density is the* TN(α) *density.*

Property 3.5 *The* DSN$(0, \alpha, \beta)$ *density is the* ETN(α, β) *density.*

Property 3.6 *If* $Z \sim \mathrm{DSN}(\varepsilon, \alpha, \beta)$ *then* $-Z \sim \mathrm{DSN}(-\varepsilon, \alpha, -\beta)$.

3.3. *Tractable sub-models*

The full doubly skewed normal model (12) provides a flexible family of densities for fitting univariate data sets. It, however, involves three shape parameters and a potentially troublesome normalizing constant. In practice it will typically be appropriate to introduce, in addition, location and scale parameters. The resulting five parameter family of densities will be of the form:

$$f_{\mathrm{DSN}}(x; \xi, \omega, \varepsilon, \alpha, \beta) = 2\frac{c_\alpha}{\omega}\left[\frac{\begin{array}{c}\phi(z/(1+\varepsilon))\Phi(\alpha|z/(1+\varepsilon)|)\Phi(\beta z/(1+\varepsilon))I(z<0)\\ +\phi(z/(1-\varepsilon))\Phi(\alpha|z/(1-\varepsilon)|)\Phi(\beta z/(1-\varepsilon))I(z\geq 0)\end{array}}{1-\varepsilon(1-2d_{\alpha,\beta})}\right]. \tag{15}$$

where $z = \omega^{-1}(x - \xi)$ and we write $X \sim \mathrm{DSN}(\boldsymbol{\theta})$ or $X \sim \mathrm{DSN}(\xi, \omega, \varepsilon, \alpha, \beta)$. The complexity of this model motivates consideration of sub-models in which one of the three parameters is set equal to zero. In some cases the full model (15) will need to be used, but it is to be hoped that frequently one of the three simpler sub-models will suffice.

The three sub-models are

(i) $f_{\mathrm{DSN}}(x; \xi, \omega, 0, \alpha, \beta)$. This is the extended two-piece skew normal density. This model is discussed in some detail in [4].
(ii) $f_{\mathrm{DSN}}(x; \xi, \omega, \varepsilon, 0, \beta)$. This is the ε-skewed version of the Azzalini model.
(iii) $f_{\mathrm{DSN}}(x; \xi, \omega, \varepsilon, \alpha, 0)$. This is the ε-skewed version of the two-piece skew normal model of Kim.

In subsequent sections, a careful analysis of sub-model (iii) will be presented. A parallel treatment of sub-model (ii) is of course possible. Sub-model (iii) is selected because it exhibits considerable flexibility and possible bi-modality (a property not shared by sub-model (ii)).

Remark 1 A natural competitor for modelling possibly bimodal data is a simple normal mixture model of the form

$$f(x; p, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2) = \frac{p}{\sigma_1}\phi\left(\frac{x-\mu_1}{\sigma_1}\right) + \frac{1-p}{\sigma_2}\phi\left(\frac{x-\mu_2}{\sigma_2}\right).$$

Such a five parameter model might be considered as an alternative in situations in which the five parameter (three shape parameters plus location and scale parameters) doubly skewed model might be applied. Typically either an E-M algorithm or a vague prior Bayesian approach using a Gibbs sampler would be used to select appropriate parameter values for fitting this model to a given data set. Both the doubly skewed model and the mixture model will require similar amounts of computational effort. Both involve five parameters. However, as can be seen from a perusal of Figure 1, the doubly skewed densities do not appear to be of the form to be expected from mixtures of normal densities. As a consequence the two models are best viewed as complementary rather than competing models. Some data sets will be better fitted by a doubly skewed model, while for other data sets a mixture model will be more appropriate.

Remark 2 [(*concomitant variables*)] In many applications, we will have, for each observation Y_i, a corresponding measurement x_i of some concomitant variable which is expected to have some influence on the variable Y_i. A regression-type model of the form $Y_i = \alpha + \beta x_i + \varepsilon_i$ might be considered in such a setting. Rather than assume that the 'error' variables, the ε_i's, are normally distributed, some authors have considered them to have a common skew normal distribution. Of course, a doubly skewed normal distribution for the ε_i's could also be considered. Such a model will have the Y_i's having a doubly skewed normal distribution with only the location parameter being affected by the concomitant variables, the x_i's. More interesting cases will arise when one allows any or all of the five parameters in the doubly skewed model to be functions of the concomitant variable, and not necessarily with a linear link function.

4. Doubly skewed model for $\boldsymbol{\beta = 0}$

The purpose of the present section is to investigate the uni/bimodal version of the doubly skewed normal density corresponding to the case in which $\beta = 0$.

In the particular case of Equation (12) in which $\beta = 0$, then the model reduces to become an ε-skewed version of the two piece normal distribution. In this case it is also possible to give an explicit expression for the normalizing constant, $d_{\alpha,0} = F(0, \alpha, 0) = 1/2$. The corresponding

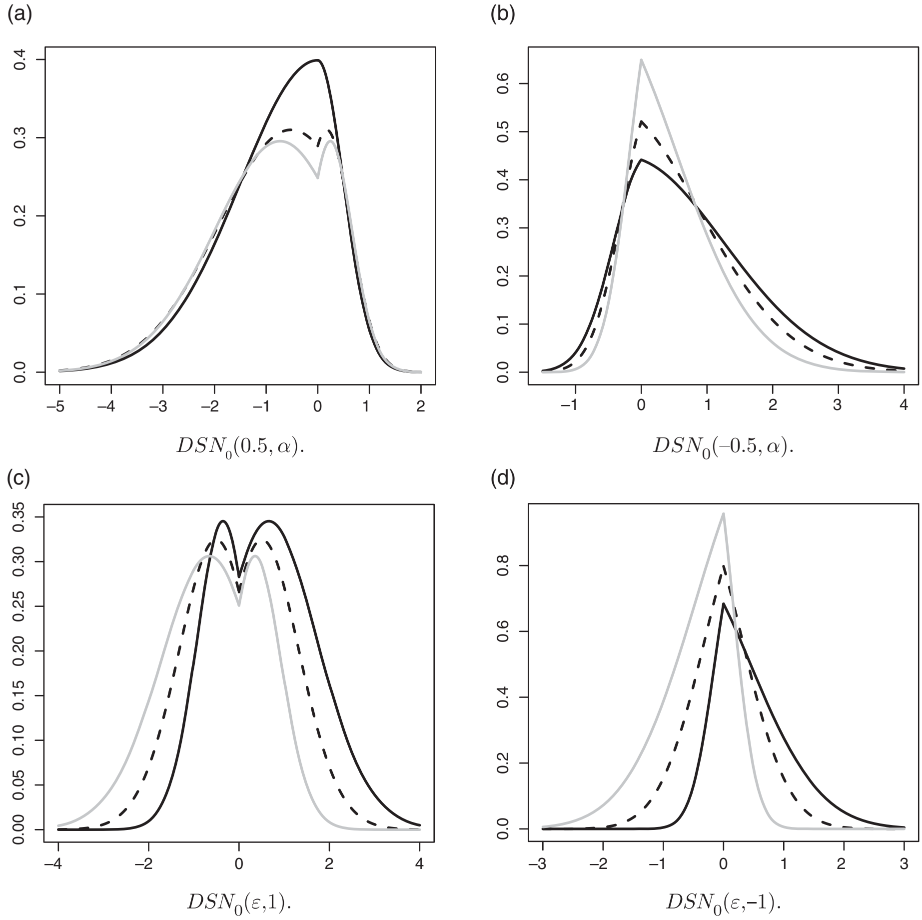

Figure 3. Examples of the DSN_0 density: (a) $\alpha = 0$ (black line), $\alpha = 0.5$ (dashed line) and $\alpha = 0.9$ (grey line); (b) $\alpha = -0.2$ (black line), $\alpha = -0.5$ (dashed line) and $\alpha = -0.9$ (grey liner); (c) $\varepsilon = -0.3$ (black line), $\varepsilon = 0$ (dashed line) and $\varepsilon = 0.3$ (grey line); (d) $\varepsilon = -0.5$ (black line), $\varepsilon = 0$ (dashed line) and $\varepsilon = 0.5$ (grey line).

density is given by

$$f_{\mathrm{DSN}}(z;\varepsilon,\alpha,0) = \begin{cases} c_\alpha \phi\left(\dfrac{z}{1+\varepsilon}\right)\left(1-\Phi\left(\dfrac{\alpha z}{1+\varepsilon}\right)\right), & z < 0, \\ c_\alpha \phi\left(\dfrac{z}{1-\varepsilon}\right)\Phi\left(\dfrac{\alpha z}{1-\varepsilon}\right), & z \geq 0. \end{cases} \tag{16}$$

Figure 3 shows the shapes of the density (16) for different parameter values. In the remainder of this paper we will refer to the $\mathrm{DSN}(\varepsilon, \alpha, 0)$ distribution as the $\mathrm{DSN}_0(\varepsilon, \alpha)$ distribution, or more briefly as the DSN_0 distribution.

4.1. *Distributional properties of the DSN_0 model*

Clearly the density (16) is continuous at $z = 0$ for every α and ε, however it is not differentiable at $z = 0$ when $\alpha \neq 0$. Below we list some properties of the DSN_0 model. Note that this model contains the normal, epsilon-skew-normal and two-piece skew-normal densities as special cases.

Property 4.1 *As $\varepsilon \to -1$, the DSN_0 distribution tends to the $c_\alpha\phi(z/2)\Phi(\alpha z/2)I(z \geq 0)$ density. In contrast, as $\varepsilon \to 1$, DSN_0 tends to the $c_\alpha\phi(z/2)\Phi(-\alpha z/2)I(z < 0)$ density.*

Property 4.2 *For $\alpha > 0$, the density (16) is bimodal, i.e. in each region of $z \in (-\infty, 0]$ and $z \in [0, \infty)$, $\log f(z; \varepsilon, \alpha, 0)$ is a concave function of z.*

Property 4.3 *For $\alpha > 0$, two modes of Equation (16) are located at $z = z_1$ and $z = z_2$ where*

$$z_1 = -\alpha(1+\varepsilon)\frac{\phi(\alpha z_1/(1+\varepsilon))}{\Phi(-\alpha z_1/(1+\varepsilon))} \quad \text{and} \quad z_2 = \alpha(1-\varepsilon)\frac{\phi(\alpha z_2/(1-\varepsilon))}{\Phi(\alpha z_2/(1-\varepsilon))}.$$

Here, $z_1 < 0$ and $z_2 > 0$.

Property 4.4 *For $\alpha < 0$, the single mode of Equation (16) is located at $z = 0$, because $f'(z; \varepsilon, \alpha, 0) < 0$ for $z > 0$ and $f'(z; \varepsilon, \alpha, 0) > 0$ for $z < 0$.*

4.2. *Distribution function*

Denote by $F_{\mathrm{DSN}_0}(z; \varepsilon, \alpha)$ the distribution function of Equation (16), i.e.

$$F_{\mathrm{DSN}_0}(z; \varepsilon, \alpha) = \begin{cases} c_\alpha \dfrac{(1+\varepsilon)}{2}\left[2\Phi\left(\dfrac{z}{1+\varepsilon}\right) - \Phi_{\mathrm{SN}}\left(\dfrac{z}{1+\varepsilon}\right)\right], & z < 0, \\ c_\alpha \dfrac{(1-\varepsilon)}{2}\left[\dfrac{\arctan(\alpha)}{\pi} - \dfrac{1}{2} + \Phi_{\mathrm{SN}}\left(\dfrac{z}{1-\varepsilon}\right)\right] + \dfrac{1+\varepsilon}{2}, & z \geq 0, \end{cases} \tag{17}$$

where Φ_{SN} is cdf of the $\mathrm{SN}(\alpha)$ distribution.

Property 4.5

$$F_{\mathrm{DSN}_0}(z; \varepsilon, 1) = \begin{cases} \dfrac{2(1+\varepsilon)}{3}\left[1 - \Phi^2\left(-\dfrac{z}{1+\varepsilon}\right)\right], & z < 0, \\ \dfrac{(1-\varepsilon)}{3}\left[1 + 2\Phi^2\left(\dfrac{z}{1-\varepsilon}\right)\right] + \varepsilon, & z \geq 0, \end{cases}$$

4.3. *Stochastic representation*

First we present two results which will be used later.

Proposition 4.1 *If $X \sim \mathrm{ETN}(\alpha, \beta)$ then the random variable $Y = |X|$ has a density of the form*:

$$f_Y(y; \alpha) = 2c_\alpha\phi(y)\Phi(\alpha y), \quad y \geq 0, \tag{18}$$

If Y thas the density given in Equation (18) we say that it is a truncated skew-normal variable and write $Y \sim \mathrm{TSN}(\alpha)$.

Corollary 1 *If $X \sim \mathrm{ETN}(\alpha, \beta)$ and $Y \sim \mathrm{TN}(\alpha)$, then $|X| \stackrel{\mathcal{D}}{=} |Y| \sim \mathrm{TSN}(\alpha)$.*

A random variable Z with density function (16) can be represented as a product of two independent random variables. This result may be proved in a manner analogous to that provided by Arellano-Valle et al. [6] in a related context. For simulation purposes, such a stochastic representation of an $\mathrm{DSN}_0(\varepsilon, \alpha)$ random variable will clearly be useful.

Proposition 4.2 *For any $\alpha \in \mathbb{R}$ and $|\varepsilon| < 1$, it follows that $Z \sim \mathrm{DSN}_0(\varepsilon, \alpha)$ if and only if there are two independent random variables Y and S with $Y \sim \mathrm{TSN}(\alpha)$ and $P(S = 1 - \varepsilon) = 1 - P(S = -(1 + \varepsilon)) = (1 - \varepsilon)/2$, such that $Z \stackrel{d}{=} SY$.*

For applications it will be convenient to add location and scale parameters to our DSN_0 distribution. If $Z \sim \mathrm{DSN}_0(\varepsilon, \alpha)$ and if $X = \xi + \omega Z$, where $\xi \in \mathbb{R}$ and $\omega > 0$, then we will write $X \sim \mathrm{DSN}_0(\xi, \omega, \varepsilon, \alpha)$ or at times $X \sim \mathrm{DSN}_0(\boldsymbol{\theta})$ where $\boldsymbol{\theta}$ denotes the vector of parameters $(\xi, \omega, \varepsilon, \alpha)$. This leads to the following definition.

Definition 4.1 *A random variable X has a distribution in the DSN_0 location and scale family if the density is given by*

$$f_{\mathrm{DSN}_0}(x; \boldsymbol{\theta}) = \begin{cases} \dfrac{c_\alpha}{\omega}\phi\left(\dfrac{z}{1+\varepsilon}\right)\left(1 - \Phi\left(\dfrac{\alpha z}{1+\varepsilon}\right)\right), & x < \xi, \\ \dfrac{c_\alpha}{\omega}\phi\left(\dfrac{z}{1-\varepsilon}\right)\Phi\left(\dfrac{\alpha z}{1-\varepsilon}\right), & x \geq \xi. \end{cases} \tag{19}$$

where $z = \omega^{-1}(x - \xi)$ and we write $X \sim \mathrm{DSN}_0(\boldsymbol{\theta})$ or $X \sim \mathrm{DSN}_0(\xi, \omega, \varepsilon, \alpha)$.

4.4. *Moments*

The following result provides a recursive formula for the moments of a random variable with density (18) which will be useful for calculating the moments of a DSN_0 random variable.

Proposition 4.3 *Let $Y \sim \mathrm{TSN}(\alpha)$, then*

$$d_r(\alpha) := E(Y^r) = \begin{cases} 1, & r = 0, \\ \dfrac{c_\alpha}{\sqrt{2\pi}}\left(1 + \dfrac{\alpha}{\sqrt{1+\alpha^2}}\right), & r = 1, \\ (r-1)d_{r-2}(\alpha) + \dfrac{2^{r/2-1}\alpha c_\alpha}{\pi(1+\alpha^2)^{r/2}}\Gamma\left(\dfrac{r}{2}\right), & r \geq 2. \end{cases} \tag{20}$$

Proof The result is achieved by integrating the expression

$$\frac{\mathrm{d}}{\mathrm{d}t}\{2t^{r-1}\phi(t)\Phi(\alpha t)\} = 2(r-1)t^{r-2}\phi(t)\Phi(\alpha t) - 2t^r\phi(t)\Phi(\alpha t) + \sqrt{2/\pi}\,\alpha t^{r-1}\phi(t\sqrt{1+\alpha^2})$$

between 0 and ∞. ■

Proposition 4.4 *Let $Z \sim \mathrm{DSN}_0(\varepsilon, \alpha)$ and $X = \xi + \omega Z \sim \mathrm{DSN}_0(\boldsymbol{\theta})$, then, for $r = 1, 2, \ldots,$ we have*

$$E(Z^r) = \frac{[(1-\varepsilon)^{r+1} + (-1)^r(1+\varepsilon)^{r+1}]d_r(\alpha)}{2} \quad \text{and} \quad E(X^r) = \sum_{j=0}^{r}\binom{r}{j}\xi^{r-j}\omega^j E(Z^j). \tag{21}$$

where $d_r(\alpha)$ is given in Equation (20).

Proof The result is readily verified using the stochastic representation given in Proposition 4.2. ■

For reference we list the first four moments of the standard DSN_0 distribution. If $Z \sim \mathrm{DSN}_0(\varepsilon, \alpha)$ then

$$E(Z) = -\sqrt{\frac{2}{\pi}}\varepsilon c_\alpha \left(1 + \frac{\alpha}{\sqrt{1+\alpha^2}}\right), \tag{22}$$

$$E(Z^2) = (1+3\varepsilon^2)\left(1 + \frac{\alpha c_\alpha}{\pi(1+\alpha^2)}\right), \tag{23}$$

$$E(Z^3) = -2\sqrt{\frac{2}{\pi}}\varepsilon(1+\varepsilon^2)c_\alpha \left(2 + \frac{\alpha(3+2\alpha^2)}{(1+\alpha^2)^{3/2}}\right), \tag{24}$$

$$E(Z^4) = (1+10\varepsilon^2+5\varepsilon^4)\left(3 + \frac{\alpha(5+3\alpha^2)c_\alpha}{\pi(1+\alpha^2)^2}\right). \tag{25}$$

Standard expressions for kurtosis and skewness can then be obtained using Equations (22)–(25).

5. Maximum likelihood estimation

5.1. *Likelihood score functions*

Let $X_1, \ldots, X_n$ be a random sample drawn from the $\mathrm{DSN}_0(\xi, \omega, \varepsilon, \alpha)$ distribution. The log-likelihood function for $\boldsymbol{\theta}$ is $\sum_{i=1}^n l(\boldsymbol{\theta}, X_i)$, where $l(\boldsymbol{\theta}, X)$ is the log-likelihood for $\boldsymbol{\theta}$ based on a single observation X, that is,

$$\begin{aligned} l(\boldsymbol{\theta}; X)) \propto \log\left(\frac{c_\alpha}{\omega}\right) - \frac{1}{2}\left(\frac{Z}{1+\varepsilon}\right)^2 I(X<\xi) + \log\Phi\left(-\frac{\alpha Z}{1+\varepsilon}\right) I(X<\xi) \\ - \frac{1}{2}\left(\frac{Z}{1-\varepsilon}\right)^2 I(X \geq \xi) + \log\Phi\left(\frac{\alpha Z}{1-\varepsilon}\right) I(X \geq \xi). \end{aligned}$$

where $Z = \omega^{-1}(X-\xi)$. The score function is $\sum_{i=1}^n S_\theta(\boldsymbol{\theta}, X_i)$, where $S_\theta(\boldsymbol{\theta}, X) = \partial l(\boldsymbol{\theta}, X)/\partial\theta$ is the vector $(S_\xi, S_\omega, S_\alpha, S_\varepsilon)$ with elements

$$\begin{aligned} S_\xi &= \left[\frac{Z}{\omega(1+\varepsilon)^2} + \frac{\alpha}{\omega(1+\varepsilon)}R(Z)\right] I(X<\xi) + \left[\frac{Z}{\omega(1-\varepsilon)^2} - \frac{\alpha}{\omega(1-\varepsilon)}S(Z)\right] I(X \geq \xi), \\ S_\omega &= -\frac{1}{\omega} + \left[\frac{Z^2}{\omega(1+\varepsilon)^2} + \frac{\alpha Z R(Z)}{\omega(1+\varepsilon)}\right] I(X<\xi) + \left[\frac{Z^2}{\omega(1-\varepsilon)^2} - \frac{\alpha Z S(Z)}{\omega(1-\varepsilon)}\right] I(X \geq \xi), \\ S_\alpha &= -\frac{c_\alpha}{\pi(1+\alpha^2)} - \frac{Z}{1+\varepsilon}R(Z)\, I(X<\xi) + \frac{Z}{1-\varepsilon}S(Z)\, I(X \geq \xi), \\ S_\varepsilon &= \left[\frac{Z^2}{(1+\varepsilon)^3} + \frac{\alpha Z}{(1+\varepsilon)^2}R(Z)\right] I(X<\xi) + \left[-\frac{Z^2}{(1-\varepsilon)^3} + \frac{\alpha Z}{(1-\varepsilon)^2}S(Z)\right] I(X \geq \xi), \end{aligned}$$

where $R(Z) = \phi(\alpha Z/(1+\varepsilon))/\Phi(-\alpha Z/(1+\varepsilon))$ and $S(Z) = \phi(\alpha Z/(1-\varepsilon))/\Phi(\alpha Z/(1-\varepsilon))$.

5.2. *The information matrix*

For one observation $X \sim \mathrm{DSN}_0(\boldsymbol{\theta})$, the i, jth element of the information matrix I is given by

$$I_{\theta_i\theta_j} = E\left[-\frac{\partial^2 l(\boldsymbol{\theta}; X)}{\partial\theta_i\partial\theta_j}\right].$$

The corresponding second partial derivatives of the log-likelihood are listed in the appendix. Eventually, one obtains the following expressions for the elements of the information matrix:

$$\begin{aligned}
I_{\xi\xi} &= \frac{1}{\omega^2(1-\varepsilon^2)} + \frac{\alpha^3 c_\alpha}{\pi\omega^2(1-\varepsilon^2)(1+\alpha^2)} + \frac{2\alpha^2 c_\alpha}{\omega^2(1-\varepsilon^2)}\rho_0(\alpha),\\
I_{\xi\varepsilon} &= \frac{\sqrt{2}c_\alpha}{\sqrt{\pi}\omega(\varepsilon^2-1)} + \frac{\sqrt{2}\alpha(1+2\alpha^2)c_\alpha}{2\sqrt{\pi}\omega(\varepsilon^2-1)(1+\alpha^2)^{3/2}} + \frac{2\alpha^2 c_\alpha}{\omega(\varepsilon^2-1)}\rho_1(\alpha),\\
I_{\omega\omega} &= \frac{2}{\omega^2} + \frac{\alpha(1+3\alpha^2)c_\alpha}{\pi\omega^2(1+\alpha^2)^2} + \frac{2\alpha^2 c_\alpha}{\omega^2}\rho_2(\alpha),\\
I_{\omega\alpha} &= \frac{(1-2\alpha^2)c_\alpha}{\pi\omega(1+\alpha^2)^2} - \frac{2\alpha c_\alpha}{\omega}\rho_2(\alpha),\\
I_{\alpha\alpha} &= 2c_\alpha\rho_2(\alpha) - \frac{c_\alpha^2}{\pi^2(1+\alpha^2)^2},\\
I_{\varepsilon\varepsilon} &= \frac{3}{1-\varepsilon^2} + \frac{\alpha(1+3\alpha^2)c_\alpha}{\pi(1-\varepsilon^2)(1+\alpha^2)^2} + \frac{2\alpha^2 c_\alpha}{1-\varepsilon^2}\rho_2(\alpha),\\
I_{\xi\omega} &= I_{\xi\alpha} = I_{\omega\varepsilon} = I_{\alpha\varepsilon} = 0,
\end{aligned}$$

where

$$\rho_r(\alpha) := \int_0^\infty t^r \frac{\phi^2(\alpha t)}{\Phi(\alpha t)}\phi(t)\,\mathrm{d}t, \quad r = 0, 1, 2. \tag{26}$$

which have to be evaluated numerically.

5.2.1. *Special cases*

The important point is that there is sufficient regularity for the asymptotic normality of the MLEs to hold when I is non-singular, and in this case, they have covariance matrix I^{-1}. Here, I is non-singular and the parameters ω and α are orthogonal to ξ and to ε. The elements of the two-piece skew-normal case are obtained by replacing $\varepsilon = 0$. In the epsilon-skew-normal case for which $\alpha = 0$, the non-zero elements of the information matrix are

$$I_{\xi\xi} = \frac{1}{\omega^2(1-\varepsilon^2)}, \quad I_{\xi\varepsilon} = \frac{2\sqrt{2}}{\sqrt{\pi}\omega(\varepsilon^2-1)}, \quad I_{\omega\omega} = \frac{2}{\omega^2}, \quad I_{\omega\alpha} = \frac{2}{\pi\omega},$$
$$I_{\alpha\alpha} = \frac{2\pi-4}{\pi^2}, \quad I_{\varepsilon\varepsilon} = \frac{3}{1-\varepsilon^2}.$$

In the normal case for which $(\varepsilon, \alpha) = (0, 0)$, the non-zero elements of the information matrix are

$$I_{\xi\xi} = \frac{1}{\omega^2}, \quad I_{\xi\varepsilon} = -\frac{2\sqrt{2}}{\sqrt{\pi}\omega}, \quad I_{\omega\omega} = \frac{2}{\omega^2}, \quad I_{\omega\alpha} = \frac{2}{\pi\omega}, \quad I_{\alpha\alpha} = \frac{2\pi-4}{\pi^2}, \quad I_{\varepsilon\varepsilon} = 3.$$

Remark 3 Gómez et al. [11] introduced an alternative possibly bimodal extension of the skew-normal distribution, which they called a skew-flexible-normal distribution. The corresponding density is of the form

$$f_{\mathrm{SFN}}(z;\alpha,\delta) = \frac{\phi(|z|+\delta)\Phi(\alpha z)}{1-\Phi(\delta)}, \quad z \in (-\infty,\infty), \tag{27}$$

where $\alpha, \delta \in (-\infty, \infty)$. This density is bimodal if $\delta < 0$.

Although both models (19) and (27) yield possibly bimodal extensions of the skew-normal distribution, their genesis is different and they differ considerably in their properties. One major difference between the two models is that, although both include the classical normal distribution as a special case, the Fisher information matrix for Equation (27) is not always non-singular as contrasted with the situation for the model (19), for which the Fisher information matrix is always non-singular.

6. Two illustrative data sets

To illustrate the estimation procedure discussed in the previous section we will use two data sets. We consider the variables *N-Cream* (first example) and *E-Shiny* (second example) available in the database creaminess of cream cheese which can be found at http://www.models.kvl.dk/Cream. Table 1 shows summary statistics for these two examples.

In Tables 2 and 3, three models are fitted to the data in the first and second example, respectively. They are $\mathrm{ESN}(\varepsilon) = \mathrm{DSN}_0(0,\varepsilon)$, $\mathrm{TN}(\alpha) = \mathrm{DSN}_0(\alpha,0)$ and $\mathrm{DSN}_0(\varepsilon,\alpha)$. In all cases, the models are augmented by the inclusion of location (ξ) and scale (ω) parameters.

In all cases, the parameters are estimated by maximum likelihood using the `bbmle` package for `R`.[12] The standard errors of the maximum likelihood estimates are calculated using the information matrix corresponding to each model.

To compare the DSN_0 model with the ESN and TN models for the first data set, consider testing the null hypothesis of an ESN or a TN distribution against a DSN_0 distribution using the likelihood ratio statistics based on the ratios $\Lambda_1 = L_{\mathrm{ESN}}(\hat{\mu}, \hat{\sigma}, \hat{\varepsilon})/$

Table 1. Descriptive statistics for the first and second examples.

Example	n	$\bar{x}$	s	γ_1	γ_2
1	240	7.578	1.716	−0.551	0.173
2	240	8.109	2.434	−0.209	−0.466

Note: Here, γ_1 and γ_2 denote the sample skewness and kurtosis coefficients, respectively.

Table 2. Estimated parameters and log-likelihood for the models ESN, TN and DSN_0 for the first example.

MLEs	ESN	TN	DSN_0
ξ	8.348(0.263)	7.650(0.812)	8.400(0.191)
ω	1.664(0.216)	2.010(0.212)	6.084(0.345)
α		−0.378(0.202)	−2.846(0.178)
ε	0.302(0.091)		0.320(0.052)
Log-likelihood	−462.696	−469.419	−459.956

Note: The respective standard errors are in parentheses.

Table 3. Estimated parameters and log-likelihood for the models Normal, ESN, TN and DSN_0 for the second example.

MLEs	N	ESN	TN	DSN_0
ξ	8.109(0.403)	9.048(0.390)	8.148(0.143)	8.907(0.165)
ω	2.429(0.315)	2.413(0.313)	2.265(0.087)	2.210(0.077)
α			1.986(1.121)	1.322(0.802)
ε		0.234(0.093)		0.202(0.050)
Log-likelihood	−553.505	−552.000	−549.393	−548.533

Note: The respective standard errors are in parentheses.

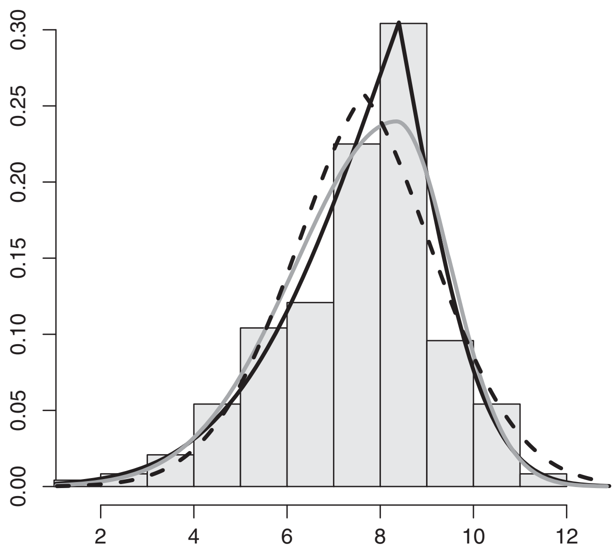

Figure 4. Histogram for the *N-Cream* variable. The curves represent densities fitted by maximum likelihood: $DSN_0(\hat{\xi}, \hat{\omega}, \hat{\varepsilon}, \hat{\alpha})$ (black line), $ESN(\hat{\xi}, \hat{\omega}, \hat{\varepsilon})$ (grey line) and $TN(\hat{\xi}, \hat{\omega}, \hat{\alpha})$ (dashed line).

$L_{DSN_0}(\hat{\mu}, \hat{\sigma}, \hat{\varepsilon}, \hat{\alpha})$ and $\Lambda_2 = L_{TN}(\hat{\mu}, \hat{\sigma}, \hat{\alpha})/L_{DSN_0}(\hat{\mu}, \hat{\sigma}, \hat{\varepsilon}, \hat{\alpha})$. Substituting the estimated values, we obtain $-2\log L(\Lambda_1) = -2(-462.696 + 459.956) = 5.48$ and $-2\log L(\Lambda_2) = -2(-469.419 + 459.956) = 18.93$ which, when compared with the 95% critical value of the $\chi_1^2 = 3.84$, indicate that the null hypotheses are clearly rejected and there is a strong indication that the DSN_0 distribution presents a much better fit than either the ESN or the TN distribution to the data set under consideration.

The summaries provided by Table 1 serve to illustrate a key feature of the DSN_0 model; its flexibility and ability to adapt to a wide range of coefficients of skewness and kurtosis, in contrast to many other models.

For the second example, we illustrate the fit of the DSN_0 model to a set of data that is more or less symmetric to see whether the parameter ε is significantly different from 0. We use data on the variable E-Shiny, Example 2, and fit four competing models: normal, epsilon-skew-normal, two-piece skew-normal and DSN_0 (Figure 5). However a 95% confidence interval for ε is given by (0.103,0.301) which does not include 0 and which indicates that, for this data set, ε is significantly different from 0. Consequently, for this data set, as for the first data set, the DSN_0 model appears to be a good choice to fit the data.

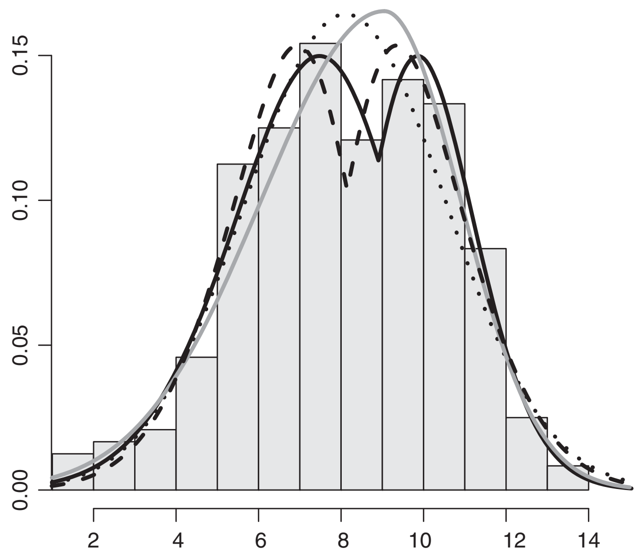

Figure 5. Histogram for the *E-Shiny* variable. The curves represent densities fitted by maximum likelihood: $\mathrm{DSN}_0(\hat{\xi}, \hat{\omega}, \hat{\varepsilon}, \hat{\alpha})$(black line), $\mathrm{ESN}(\hat{\xi}, \hat{\omega}, \hat{\varepsilon})$ (grey line), $\mathrm{TN}(\hat{\xi}, \hat{\omega}, \hat{\alpha})$ (dashed line) and $N(\hat{\xi}, \hat{\omega})$ (dotted line).

The conclusion that the DSN_0 model appears to be more appropriate for both of the data sets analysed here is supported by inspection of Figures 4 and 5 where the histograms and the fitted curves for the data sets are displayed.

Acknowledgements

The authors acknowledge helpful comments and suggestions from the referee which substantially improved the presentation. The research of H.S. Salinas was supported by DIUDA 221229. The research of H.W. Gómez was supported by FONDECYT 1130495.

References

[1] Azzalini A. A class of distributions which includes the normal ones. Scand J Stat. 1985;12:171–178.
[2] Arnold BC, Beaver RJ. Hidden truncation models. Sankhya. 2000;A62:23–35.
[3] Kim HJ. On a class of two-piece skew-normal distributions. Stat: J Theor Appl Stat. 2005;39(6):537–553.
[4] Arnold BC, Gómez HW, Salinas HS. On multiple constraint skewed models. Stat: J Theor Appl Stat. 2009;43(3):279–293.
[5] Mudholkar GS, Hutson AD. The epsilon-skew-normal distribution for analyzing near-normal data. J Statist Plann Inference. 2000;83:291–309.
[6] Arellano-Valle RB, Gómez HW, Quintana FA. Statistical inference for a general class of asymmetric distributions. J Statist Plann Inference. 2005;128(2):427–443.
[7] Arellano-Valle RB, Cortés MA, Gómez HW. An extension of the epsilon-skew-normal distribution. Commun Stat: Theory Methods. 2010;39:912–922.
[8] Pewsey A, Gómez HW, Bolfarine H. Likelihood-based inference for power distributions. Test. 2012;21(4):775–789.
[9] Gómez HW, Torres FJ, Bolfarine H. Large-sample inference for the epsilon-skew-t distribution. Commun Stat. Theory Methods. 2007;36(1):73–81.
[10] Gómez HW, Salinas HS, Bolfarine H. Generalized skew-normal models: properties and inference. Stat: J Theor Appl Stat. 2006;40(6):495–505.
[11] Gómez HW, Elal-Olivero D, Salinas HS, Bolfarine H. Bimodal extension based on the skew-normal distribution with application to pollen data. Environmetrics. 2011;22(1):50–62.
[12] R Development Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2011. ISBN 3-900051-07-0. Available from: http://www.R-project.org

Appendix

(a) The second-order derivatives of $l(\boldsymbol{\theta};X)$ are

$$
\begin{aligned}
\frac{\partial^2 l(\boldsymbol{\theta};X)}{\partial\xi^2} &= -\frac{1}{\omega^2}\left[\frac{1}{(1+\varepsilon)^2} - \frac{\alpha^3 Z}{(1+\varepsilon)^3}R(Z) + \frac{\alpha^2}{(1+\varepsilon)^2}R^2(Z)\right] I(X<\xi)\\
&\quad - \frac{1}{\omega^2}\left[\frac{1}{(1-\varepsilon)^2} + \frac{\alpha^3 Z}{(1-\varepsilon)^3}S(Z) + \frac{\alpha^2}{(1-\varepsilon)^2}S^2(Z)\right] I(X\geq\xi),\\
\frac{\partial^2 l(\boldsymbol{\theta};X)}{\partial\omega\partial\xi} &= -\frac{1}{\omega^2}\left[\frac{2Z}{(1+\varepsilon)^2} + \frac{\alpha}{(1+\varepsilon)}R(Z) - \frac{\alpha^3 Z^2}{(1+\varepsilon)^3}R(Z) + \frac{\alpha^2 Z}{(1+\varepsilon)^2}R^2(Z)\right] I(X<\xi)\\
&\quad - \frac{1}{\omega^2}\left[\frac{2Z}{(1-\varepsilon)^2} - \frac{\alpha}{(1-\varepsilon)}S(Z) + \frac{\alpha^3 Z^2}{(1-\varepsilon)^3}S(Z) + \frac{\alpha^2 Z}{(1-\varepsilon)^2}S^2(Z)\right] I(X\geq\xi),\\
\frac{\partial^2 l(\boldsymbol{\theta};X)}{\partial\alpha\partial\xi} &= -\frac{1}{\omega}\left[-\frac{1}{1+\varepsilon}R(Z) + \frac{\alpha^2 Z^2}{(1+\varepsilon)^3}R(Z) - \frac{\alpha Z}{(1+\varepsilon)^2}R^2(Z)\right] I(X<\xi)\\
&\quad - \frac{1}{\omega}\left[\frac{1}{1-\varepsilon}S(Z) - \frac{\alpha^2 Z^2}{(1-\varepsilon)^3}S(Z) - \frac{\alpha Z}{(1-\varepsilon)^2}S^2(Z)\right] I(X\geq\xi),\\
\frac{\partial^2 l(\boldsymbol{\theta};X)}{\partial\varepsilon\partial\xi} &= -\frac{1}{\omega}\left[\frac{2Z}{(1+\varepsilon)^3} + \frac{\alpha}{(1+\varepsilon)^2}R(Z) - \frac{\alpha^3 Z^2}{(1+\varepsilon)^4}R(Z) + \frac{\alpha^2 Z}{(1+\varepsilon)^3}R^2(Z)\right] I(X<\xi)\\
&\quad - \frac{1}{\omega}\left[-\frac{2Z}{(1-\varepsilon)^3} + \frac{\alpha}{(1-\varepsilon)^2}S(Z) - \frac{\alpha^3 Z^2}{(1-\varepsilon)^4}S(Z) - \frac{\alpha^2 Z}{(1-\varepsilon)^3}S^2(Z)\right] I(X\geq\xi),\\
\frac{\partial^2 l(\boldsymbol{\theta};X)}{\partial\omega^2} &= \frac{1}{\omega^2} - \frac{1}{\omega^2}\left[\frac{3Z^2}{(1+\varepsilon)^2} + \frac{2\alpha Z}{1+\varepsilon}R(Z) - \frac{\alpha^3 Z^3}{(1+\varepsilon)^3}R(Z) + \frac{\alpha^2 Z^2}{(1+\varepsilon)^2}R^2(Z)\right] I(X<\xi)\\
&\quad - \frac{1}{\omega^2}\left[\frac{3Z^2}{(1-\varepsilon)^2} - \frac{2\alpha Z}{1-\varepsilon}S(Z) + \frac{\alpha^3 Z^3}{(1-\varepsilon)^3}S(Z) + \frac{\alpha^2 Z^2}{(1-\varepsilon)^2}S^2(Z)\right] I(X\geq\xi),\\
\frac{\partial^2 l(\boldsymbol{\theta};X)}{\partial\alpha\partial\omega} &= -\frac{1}{\omega}\left[-\frac{Z}{1+\varepsilon}R(Z) + \frac{\alpha^2 Z^3}{(1+\varepsilon)^3}R(Z) - \frac{\alpha Z^2}{(1+\varepsilon)^2}R^2(Z)\right] I(X<\xi)\\
&\quad - \frac{1}{\omega}\left[\frac{Z}{1-\varepsilon}S(Z) - \frac{\alpha^2 Z^3}{(1-\varepsilon)^3}S(Z) - \frac{\alpha Z^2}{(1-\varepsilon)^2}S^2(Z)\right] I(X\geq\xi),\\
\frac{\partial^2 l(\boldsymbol{\theta};X)}{\partial\varepsilon\partial\omega} &= -\frac{1}{\omega}\left[\frac{2Z^2}{(1+\varepsilon)^3} + \frac{\alpha Z}{(1+\varepsilon)^2}R(Z) - \frac{\alpha^3 Z^3}{(1+\varepsilon)^4}R(Z) + \frac{\alpha^2 Z^2}{(1+\varepsilon)^3}R^2(Z)\right] I(X<\xi)\\
&\quad - \frac{1}{\omega}\left[-\frac{2Z^2}{(1-\varepsilon)^3} + \frac{\alpha Z}{(1-\varepsilon)^2}S(Z) - \frac{\alpha^3 Z^3}{(1-\varepsilon)^4}S(Z) - \frac{\alpha^2 Z^2}{(1-\varepsilon)^3}S^2(Z)\right] I(X\geq\xi),\\
\frac{\partial^2 l(\boldsymbol{\theta};X)}{\partial\alpha^2} &= \frac{c_\alpha(c_\alpha+2\pi\alpha)}{\pi^2(1+\alpha^2)^2} - \left[-\frac{\alpha Z^3}{(1+\varepsilon)^3}R(Z) + \frac{Z^2}{(1+\varepsilon)^2}R^2(Z)\right] I(X<\xi)\\
&\quad - \left[\frac{\alpha Z^3}{(1-\varepsilon)^3}S(Z) + \frac{Z^2}{(1-\varepsilon)^2}S^2(Z)\right] I(X\geq\xi),\\
\frac{\partial^2 l(\boldsymbol{\theta};X)}{\partial\varepsilon\partial\alpha} &= -\left[-\frac{Z}{(1+\varepsilon)^2}R(Z) + \frac{\alpha^2 Z^3}{(1+\varepsilon)^4}R(Z) - \frac{\alpha Z^2}{(1+\varepsilon)^3}R^2(Z)\right] I(X<\xi)\\
&\quad - \left[-\frac{Z}{(1-\varepsilon)^2}S(Z) + \frac{\alpha^2 Z^3}{(1-\varepsilon)^4}S(Z) + \frac{\alpha Z^2}{(1-\varepsilon)^3}S^2(Z)\right] I(X\geq\xi)
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial^2 l(\boldsymbol{\theta};X)}{\partial\varepsilon^2} &= -\left[\frac{3Z^2}{(1+\varepsilon)^4} + \frac{2\alpha Z}{(1+\varepsilon)^3}R(Z) - \frac{\alpha^3 Z^3}{(1+\varepsilon)^5}R(Z) + \frac{\alpha^2 Z^2}{(1+\varepsilon)^4}R^2(Z)\right] I(X<\xi)\\
&\quad - \left[\frac{3Z^2}{(1-\varepsilon)^4} - \frac{2\alpha Z}{(1-\varepsilon)^3}S(Z) + \frac{\alpha^3 Z^3}{(1-\varepsilon)^5}S(Z) + \frac{\alpha^2 Z^2}{(1-\varepsilon)^4}S^2(Z)\right] I(X\geq\xi).
\end{aligned}
$$

The following proposition gives the truncated moments that are needed to calculate the information matrix of the DSN_0 model.

PROPOSITION .1 *For integer $r \geq 0$ and $Z \sim \mathrm{DSN}_0(\varepsilon, \alpha)$,*

(a) $E_{Z<0}(Z^r) = (-1)^r(1+\varepsilon)^{r+1}d_r(\alpha)/2$ *and* $E_{Z\geq 0}(Z^r) = (1-\varepsilon)^{r+1}d_r(\alpha)/2$,

(b)

$$E_{Z<0}(Z^r R(Z)) = (-1)^r \frac{2^{(r-3)/2}(1+\varepsilon)^{r+1}c_\alpha}{\pi(1+\alpha^2)^{(r+1)/2}} \Gamma\left(\frac{r+1}{2}\right),$$

(c)

$$E_{Z\geq 0}(Z^r S(Z)) = \frac{2^{(r-3)/2}(1-\varepsilon)^{r+1}c_\alpha}{\pi(1+\alpha^2)^{(r+1)/2}} \Gamma\left(\frac{r+1}{2}\right),$$

(d) $E_{Z<0}(Z^r R^2(Z)) = (-1)^r(1+\varepsilon)^{r+1}c_\alpha\rho_r(\alpha)$ *and* $E_{Z\geq 0}(Z^r S^2(Z)) = (1-\varepsilon)^{r+1}c_\alpha\rho_r(\alpha)$,

where $d_r(\alpha)$ and $\rho_r(\alpha)$ are given in Equations (20) and (26), respectively.

Proof Applying the definition of mathematical expectation and making changes of variables, $t = -z/(1+\varepsilon)$ and $u = z/(1-\varepsilon)$, respectively, the results follow. ■

(b) Proof of Proposition 4.1.

Proof Let $X \sim \mathrm{ETN}(\alpha, \beta)$ then the density function of $Y = |X|$ is given by

$$\begin{aligned} f_Y(y) &= f_X(y; \alpha, \beta) + f_X(-y; \alpha, \beta) \\ &= 2c_\alpha\phi(y)\Phi(\alpha|y|)\Phi(\beta y) + 2c_\alpha\phi(y)\Phi(\alpha|-y|)\Phi(-\beta y) \\ &= 2c_\alpha\phi(y)\Phi(\alpha y)[\Phi(\beta y) + \Phi(-\beta y)] \\ &= 2c_\alpha\phi(y)\Phi(\alpha y)I(y \geq 0). \end{aligned}$$

■